STRAINED SILICON MOSFET HAVING SILICON SOURCE/DRAIN REGIONS AND METHOD FOR ITS FABRICATION

RELATED APPLICATIONS

[0001] This application is a divisional of copending application Serial No. 10/282,538, now issued as U.S. Patent 6,657,223, the entirety of which is incorporated herein by reference,

BACKGROUND OF THE INVENTION

[0002] Field of the Invention

[0003] The present invention relates generally to fabrication of metal oxide semiconductor field effect transistors (MOSFETs), and, more particularly, to MOSFETs that achieve improved carrier mobility through the incorporation of strained silicon.

[0004] Related technology

[0005] MOSFETs are a common component of integrated circuits (ICs). Figure 1 shows a conventional MOSFET device. The MOSFET is fabricated on a semiconductor substrate 10 within an active area bounded by shallow trench isolations 12 that electrically isolate the active area of the MOSFET from other IC components fabricated on the substrate 10.

[0006] The MOSFET is comprised of a gate electrode 14 that is separated from a channel region 16 in the substrate 10 by a thin first gate insulator 18 such as silicon oxide or oxide-nitride-oxide (ONO). To minimize the resistance of the gate 14, the gate 14 is typically formed of a doped semiconductor material such as polysilicon.

[0007] The source and drain of the MOSFET are provided as deep source and drain regions 20 formed on opposing sides of the gate 14. Source and drain silicides 22 are formed on the source and drain regions 20 and are comprised of a compound comprising the substrate semiconductor material and a metal such as cobalt (Co) or nickel (Ni) to reduce contact resistance to the source and drain regions 20. The source and drain regions 20 are formed deeply enough to extend beyond the depth to which the source and drain

silicides 22 are formed. The source and drain regions 20 are implanted subsequent to the formation of a spacer 24 around the gate 14 which serves as an implantation mask to define the lateral position of the source and drain regions 20 relative to the channel region 16 beneath the gate.

[0008] The gate 14 likewise has a silicide 26 formed on its upper surface. The gate structure comprising a polysilicon material and an overlying silicide is sometimes referred to as a polycide gate.

[0009] The source and drain of the MOSFET further comprise shallow source and drain extensions 28. As dimensions of the MOSFET are reduced, short channel effects resulting from the small distance between the source and drain cause degradation of MOSFET performance. The use of shallow source and drain extensions 28 rather than deep source and drain regions near the ends of the channel 16 helps to reduce short channel effects. The shallow source and drain extensions are implanted prior to the formation of the spacer 24 and after the formation of a thin spacer 30, and the gate 14 and thin spacer 30 act as an implantation mask to define the lateral position of the shallow source and drain extensions 28 relative to the channel region 16. Diffusion during subsequent annealing causes the source and drain extensions 28 to extend slightly beneath the gate 14.

[0010] One option for increasing the performance of MOSFETs is to enhance the carrier mobility of silicon so as to reduce resistance and power consumption and to increase drive current, frequency response and operating speed. A method of enhancing carrier mobility that has become a focus of recent attention is the use of silicon material to which a tensile strain is applied. "Strained" silicon may be formed by growing a layer of silicon on a silicon germanium substrate. The silicon germanium lattice is generally more widely spaced than a pure silicon lattice as a result of the presence of the larger germanium atoms in the lattice. Because the atoms of the silicon lattice align with the more widely spread silicon germanium lattice, a tensile strain is created in the silicon layer. The silicon atoms are essentially pulled apart from one another. The amount of tensile strain applied to the silicon lattice increases with the proportion of germanium in the silicon germanium lattice.

[0011] Relaxed silicon has six equal valence bands. The application of tensile strain to the silicon lattice causes four of the valence bands to increase in energy and two of the valence bands to decrease in energy. As a result of quantum effects, electrons effectively weigh 30 percent less when passing through the lower energy bands. Thus the lower energy bands offer less resistance to electron flow. In addition, electrons encounter less vibrational energy from the nucleus of the silicon atom, which causes them to scatter at a rate of 500 to 1000 times less than in relaxed silicon. As a result, carrier mobility is dramatically increased in strained silicon as compared to relaxed silicon, offering a potential increase in mobility of 80% or more for electrons and 20% or more for holes. The increase in mobility has been found to persist for current fields of up to 1.5 megavolts/centimeter. These factors are believed to enable a device speed increase of 35% without further reduction of device size, or a 25% reduction in power consumption without a reduction in performance.

[0012] An example of a MOSFET using a strained silicon layer is shown in Figure 2. The MOSFET is fabricated on a substrate comprising a silicon germanium layer 32 on which is formed an epitaxial layer of strained silicon 34. The MOSFET uses conventional MOSFET structures including deep source and drain regions 20, shallow source and drain extensions 28, a gate oxide layer 18, a gate 14 surrounded by spacers 30, 24, silicide source and drain contacts 22, a silicide gate contact 26, and shallow trench isolations 12. The channel region 16 of the MOSFET includes the strained silicon material, which provides enhanced carrier mobility between the source and drain.

[0013] One detrimental property of strained silicon MOSFETs of the type shown in Figure 2 is that the band gap of silicon germanium is lower than that of silicon. In other words, the amount of energy required to move an electron into the conduction band is lower on average in a silicon germanium lattice than in a silicon lattice. As a result, the junction leakage in devices having their source and drain regions formed in silicon germanium is greater than in comparable devices having their source and drain regions formed in silicon.

[0014] Another detrimental property of strained silicon MOSFETs of the type shown in Figure 2 is that the dielectric constant of silicon germanium is higher than that of silicon. As a result, MOSFETs incorporating silicon germanium exhibit higher parasitic capacitance, which increases device power consumption and decreases driving current and frequency response.

[0015] Therefore the advantages achieved by incorporating strained silicon into MOSFET designs are partly offset by the disadvantages resulting from the use of a silicon germanium substrate.

SUMMARY OF THE INVENTION

[0016] It is an object of the present invention to provide a strained silicon MOSFET device that exploits the benefits of strained silicon while reducing the detrimental effects of the use of silicon germanium to support the strained silicon layer.

[0017] In accordance with embodiments of the invention, a MOSFET incorporates a strained silicon layer that is supported by a silicon germanium layer. Strained silicon and silicon germanium at the locations of deep source and drain regions are removed and replaced with silicon regions. Deep source and drain regions are then implanted in the silicon regions. The formation of source and drain regions in the silicon regions reduces junction leakage and parasitic capacitance and therefore improves device performance compared to the conventional strained silicon MOSFET.

[0018] In accordance with one embodiment of the invention, a MOSFET incorporating strained silicon is fabricated. Initially a substrate is provided. The substrate includes a layer of silicon germanium having a layer of strained silicon formed thereon. The substrate further includes a gate insulator formed on the strained silicon layer and a gate formed on the gate insulator, and shallow source and drain extensions formed at opposing sides of the gate. A spacer is then formed around the gate and gate insulator. The strained silicon layer and silicon germanium layer are then etched to form trenches adjacent to the spacer at the opposing sides of the gate. Silicon regions are then formed in the trenches, and deep source and drain regions are formed in the silicon regions at

the opposing sides of the gate. The depth of the deep source and drain regions does not extend beyond the depth of the silicon regions.

[0019] In accordance with another embodiment of the invention, a MOSFET incorporating strained silicon is provided. The MOSFET includes a substrate comprising a layer of silicon germanium, and a gate that overlies a strained silicon layer formed on the silicon germanium layer and that is separated from the strained silicon layer by a gate insulator. Silicon regions are formed at opposing ends of the gate adjacent to ends of the strained silicon layer. Shallow source and drain extensions are formed at opposing ends of the gate in the strained silicon layer, and deep source and drain regions are formed at the opposing ends of the gate in the silicon regions. The depth of the deep source and drain regions does not extend beyond the depth of the silicon regions.

DESCRIPTION OF THE DRAWINGS

[0020] Embodiments of the invention are described in conjunction with the following drawings, in which:

[0021] Figure 1 shows a conventional MOSFET formed in accordance with conventional processing;

[0022] Figure 2 shows a strained silicon MOSFET device formed in accordance with the conventional processing used to form the MOSFET of Figure 1;

[0023] Figures 3a, 3b, 3c, 3d, 3e, 3f, 3g, 3h, 3i and 3j show structures formed during production of a MOSFET device in accordance with a first preferred embodiment of the invention;

[0024] Figure 4 shows a process flow encompassing the first preferred embodiment and alternative embodiments.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0025] Figures 3a - 3i show structures formed during fabrication of a strained silicon MOSFET in accordance with a preferred embodiment of the invention. Figure 3a shows a structure comprising a layer of silicon germanium

40 having an epitaxial layer of strained silicon 42 formed on its surface. The silicon germanium layer 40 preferably has a composition $Si_{1-x}Ge_x$, where x is approximately .2, and is more generally in the range of .1 to .3. The silicon germanium layer 40 is typically grown on a silicon wafer. Silicon germanium may be grown, for example, by chemical vapor deposition using Si_2H_6 (disilane) and GeH_4 (germane) as source gases, with a substrate temperature of 600 to 900 degrees C, a Si_2H_6 partial pressure of 30 mPa, and a GeH_4 partial pressure of 60 mPa. SiH_4 (silane) may be used in alternative processes. Growth of the silicon germanium material may be initiated using these ratios, or alternatively the partial pressure of GeH_4 may be gradually increased beginning from a lower pressure or zero pressure to form a gradient composition. The thickness of the silicon germanium layer may be determined in accordance with the particular application. The upper portion of the silicon germanium substrate 40 on which the strained silicon layer 42 is grown should have a uniform composition.

[0026] The strained silicon layer 42 is preferably grown by chemical vapor deposition (CVD) using Si_2H_6 as a source gas with a partial pressure of 30mPa and a substrate temperature of approximately 600 to 900 degrees C. The strained silicon layer is preferably grown to a thickness of 200 Angstroms.

[0027] As further shown in Figure 3a, a gate insulating layer 44 is formed on the strained silicon layer 42. The gate insulating layer 44 is typically silicon oxide but may be another material such as oxide-nitride-oxide (ONO). An oxide may be grown by thermal oxidation of the strained silicon layer, but is preferably deposited by chemical vapor deposition.

[0028] Formed over the gate insulating layer 44 is a gate conductive layer 46. The gate conductive layer 46 typically comprises polysilicon but may alternatively comprise another material such as polysilicon implanted with germanium.

[0029] Overlying the gate conductive layer 46 is a bi-layer hardmask structure comprising a bottom hardmask layer 48, also referred to as a bottom antireflective coating (BARC), and an upper hardmask layer 50. The bottom hardmask layer 48 is typically silicon oxide (e.g. SiO_2) and the upper hardmask layer 50 is typically silicon nitride (e.g. Si_3N_4).

[0030] The silicon germanium substrate also has formed therein shallow trench isolations 52. The shallow trench isolations may be formed by forming trenches having tapered sidewalls in the silicon germanium 40 and strained silicon 42 layers, performing a brief thermal oxidation, and then depositing a layer of silicon oxide to a thickness that is sufficient to fill the trenches, such as by low pressure CVD (LPCVD) TEOS or atmospheric pressure ozone TEOS. The silicon oxide layer is then densified and planarized such as by chemical mechanical polishing or an etch back process, leaving shallow trench isolations 52 that are approximately level with the surface of the strained silicon layer 42.

[0031] Figure 3b shows the structure of Figure 3a after patterning of the gate conductive layer and gate insulating layer to form a gate 54 and a self-aligned gate insulator 56. Patterning is performed using a series of anisotropic etches that patterns the upper hardmask layer using a photoresist mask as an etch mask, then patterns the lower hardmask layer using the patterned upper hardmask layer as an etch mask, then patterns the polysilicon using the patterned lower hardmask layer as an etch mask, then patterns the gate insulating layer using the gate 54 as a hardmask. As shown in Figure 3b, the thickness of the lower hardmask layer is chosen such that after patterning of the gate insulating layer, a portion of the lower hardmask layer remains on the gate as a protective cap 58.

[0032] Figure 3c shows the structure of Figure 3b after formation of a thin first spacer 60 around the gate 54, the gate insulator 56 and the protective cap 58. The thin first spacer 60 is preferably formed by deposition of a conformal layer of a protective material, followed by anisotropic etching to remove the protective material from the non-vertical surfaces to leave the thin first gate spacer 60. The thin first spacer 60 is preferably formed of silicon oxide or silicon nitride.

[0033] Figure 3d shows the structure of Figure 3c after implantation of dopant to form shallow source and drain extensions 62 in the strained silicon layer 42 and silicon germanium layer 40 at opposing sides of the gate 54. Halo regions (not shown) may be implanted prior to implantation of the shallow source and drain extensions 62. Halo regions are regions that are implanted

with a dopant that is opposite in conductivity type to the conductivity type of an adjacent region. The dopant of the halo regions retards diffusion of the dopant of the adjacent region. Halo regions are preferably implanted using a low energy at a small angle to the surface of the substrate so that the halo regions extend beneath the gate 54 to beyond the anticipated locations of the ends of the source and drain extensions 62 after annealing. The halo regions are formed at opposing sides of the channel region, and extend toward the channel region beyond the ends of the source and drain extensions to be formed.

[0034] Figure 3e shows the structure of Figure 3d after formation of a second spacer 64 around the first spacer 60 and gate 54. The second spacer 64 is preferably formed of a material such as silicon oxide or silicon nitride.

[0035] Figure 3f shows the structure of Figure 3e after anisotropic etching of the strained silicon layer 42 and the silicon germanium layer 40 to form trenches 66 at opposing sides of the second spacer 64 and gate 54. Typical etch chemistries are CF4 and HBr. The etch is essentially self-masking because the spacers 64 and protective cap 58 protect the gate structure, and the shallow trench isolations 52 define the outer boundaries of the etch. Therefore the edges of the trenches are aligned to the edges of the spacer 64 and the edges of the shallow trench isolations 52. The trenches 66 are etched to a depth that is great enough to contain deep source and drain regions that are formed in later processing.

[0036] Figure 3g shows the structure of Figure 3f after selective epitaxial growth of silicon in the trenches to form silicon regions 68 in the trenches at opposing sides of the gate 54. Preferably the selective growth of silicon is performed in a manner that produces no silicon growth on regions other than the exposed crystalline surfaces of the silicon germanium layer 40 and the strained silicon. Such growth may be performed, for example, by chemical vapor deposition using $SiBr_4$ as a source gas. Alternatively, $SiHCl_3$ may be used, or a mixture of SiH_2Cl_2, SiH_4 and HCl or Cl_2 may be used. As a general matter, the selectivity of the deposition process is improved by decreased pressure, increased temperature, and a decreased mole fraction of silicon in the source gas stream. The selective growth process produces crystalline silicon growth on

the exposed crystalline surfaces of the silicon germanium 40 and strained silicon. Any silicon material deposited on other surfaces such as the second gate spacer 64, the shallow trench isolations 52 and the gate protective cap 58 will be polycrystalline in form. Where selectivity cannot be precisely controlled, it may be desirable to follow selective growth of silicon with a brief exposure to an etchant that is highly selective to polysilicon so as to remove any unwanted polysilicon material from structures such as the gate spacer 64, the shallow trench isolations 52 and the gate protective cap 58. Appropriate masking, such as with photoresist, may be used to inhibit growth and facilitate removal in areas where silicon growth is not desired.

[0037] Figure 3h shows the structure of Figure 3g after formation of deep source and drain regions 70 in the silicon regions 68 at opposing sides of the gate by implantation of dopant. The second spacer 64 serves as an implant mask during implantation of the deep source and drain regions 70 to define the position of the source and drain regions 70 relative to the gate 54. The implantation is performed such that the depth of the deep source and drain regions 70 does not extend beyond the depth of the silicon regions 66 upon implantation, or after diffusion of dopant resulting from annealing as described below.

[0038] Figure 3i shows the structure of Figure 3h after performing rapid thermal annealing (RTA) to anneal the silicon regions 68 and the silicon germanium layer 40 and to activate the dopants implanted in the shallow source and drain extensions 62 and the deep source and drain regions 70. During annealing the implanted dopant undergoes diffusion, however the depth of the silicon regions 68 is chosen such that after annealing the depth of the deep source and drain regions 70 does not extend beyond the depth of the silicon regions 68. As a result the parasitic capacitance and junction leakage of the device are improved compared to a conventional strained silicon MOSFET having source and drain regions formed in a silicon germanium layer.

[0039] Figure 3j shows the structure of Figure 3i after removal of the protective gate cap 58 to expose the upper surface of the gate 54, followed by formation of silicide contacts 72 on the source and drain regions 70 and

formation of a silicide contact 74 on the gate 54. The silicide contacts 72, 74 are formed of a compound comprising a semiconductor material and a metal. Typically a metal such as cobalt (Co) is used, however other metals such as nickel (Ni) may also be employed. The silicide contacts are formed by depositing a thin conformal layer of the metal over the substrate, and then annealing to promote silicide formation at the points of contact between the metal and underlying semiconductor materials, followed by stripping of residual metal.

[0040] While the processing shown in Figures 3a - 3j represents a presently preferred embodiment, a variety of alternatives may be implemented. For example, in one alternative embodiment, a third spacer may be formed around the second spacer after growth of the silicon regions in the trenches and before implantation of the deep source and drain regions in the silicon regions. In contrast to the structure shown in Figure 3g, the use of a third spacer causes the lateral edges of the implanted deep source and drain regions to be located within the silicon regions rather than at the lateral junctions of the silicon regions with the silicon germanium layer. Through appropriate selection of the thickness of the third spacer, the structures of the deep source and drain regions may be controlled so that even after diffusion the deep source and drain regions do not project laterally into the silicon germanium layer.

[0041] Accordingly, a variety of embodiments in accordance with the invention may be implemented. In general terms, such embodiments encompass a MOSFET that includes a strained silicon channel region formed on a silicon germanium layer, and source and drain regions formed in silicon regions that are provided at opposing sides of the gate. The depth of the source and drain regions does not extend beyond the depth of the silicon regions, thus reducing the detrimental junction leakage and parasitic capacitance of conventional silicon germanium implementations.

[0042] Figure 4 shows a process flow encompassing the preferred embodiment of Figures 3a - 3j, the aforementioned alternatives and other alternatives. Initially a substrate is provided (80). The substrate includes a layer of silicon germanium having a layer of strained silicon formed thereon. The

substrate further includes a gate insulator formed on the strained silicon layer and a gate formed on the gate insulator, and shallow source and drain extensions. A spacer is then formed around the gate and gate insulator (82). The strained silicon layer and silicon germanium layer are then etched to form trenches at opposing sides of the gate (84). The edges of the trenches are aligned approximately with the edges of the spacer. Silicon regions are then formed in the trenches (86), and deep source and drain regions are implanted in the silicon regions (88). The depth of the deep source and drain regions does not extend beyond the depth of the silicon regions.

[0043] It will be apparent to those having ordinary skill in the art that the tasks described in the above processes are not necessarily exclusive of other tasks, but rather that further tasks may be incorporated into the above processes in accordance with the particular structures to be formed. For example, intermediate processing tasks such as formation and removal of passivation layers or protective layers between processing tasks, formation and removal of photoresist masks and other masking layers, doping and counter-doping, cleaning, planarization, and other tasks, may be performed along with the tasks specifically described above. Further, the process need not be performed on an entire substrate such as an entire wafer, but rather may be performed selectively on sections of the substrate. Thus, while the embodiments illustrated in the figures and described above are presently preferred, it should be understood that these embodiments are offered by way of example only. The invention is not limited to a particular embodiment, but extends to various modifications, combinations, and permutations that fall within the scope of the claimed inventions and their equivalents.